

A mosaic monoploid reference sequence for the highly complex genome of sugarcane

La canne à sucre (*Saccharum spp.*) est une graminée (Poaceae) tropicale majoritairement cultivée pour produire du sucre mais son imposante masse végétale est aussi convertible en bio-énergie. Les cultivars modernes de canne à sucre sont issus de croisements interspécifiques entre deux espèces polyploïdes, *Saccharum officinarum* ($2n = 8X = 80$ chromosomes) et *Saccharum spontaneum* ($2n = 5X$ à $12X = 40$ à 128 chromosomes). Le génome de ces cultivars est donc hautement polyploïde ($\sim 12X$), aneuploïde ($2n = 8x$ à $12x = 100$ à 130 chromosomes) et hétérozygote. Ces caractéristiques en font un des génomes les plus complexes parmi les plantes cultivées et l'obtention d'une séquence de référence de l'ensemble du génome avec les méthodes d'assemblage actuels représente donc un véritable challenge. Dans des études précédentes, nous avons montré que les chromosomes hom(e)ologues du génome de la canne à sucre partagent entre eux un très haut niveau de micro-colinéarité et présentent également une très bonne synténie avec le génome du sorgho. Ces résultats nous ont permis de proposer que le séquençage d'un génome de base (génome monoploïde) représenterait une bonne séquence de référence et que le sorgho pourrait nous permettre d'identifier cette partie du génome. Notre stratégie basée sur la technique de Whole Génome Profiling nous a permis d'identifier 4660 clones BAC de canne à sucre (BAC = fragment de chromosome d'environ 100 kb) recouvrant la partie riche en gène du génome du sorgho. Ces clones BAC ont été séquencés et assemblés en 10 chromosomes et correspondent à la partie riche en gène d'un génome monoploïde de canne à sucre représentant une séquence de 382 Mb. 25,316 modèles de gènes ont été prédit et 17% d'entre eux ne sont pas retrouvés en position colinéaires avec leur orthologue chez le sorgho. D'autre part, nous avons montré que les génomes des deux espèces à l'origine des cultivars modernes, *S. officinarum* et *S. spontaneum* ont un contenu en éléments transposables différent qui explique leur différente taille de génome de base. Ces deux espèces diffèrent également par quelques larges réarrangements chromosomiques qui expliquent leur nombre de chromosomes de base différents.